American Community Survey (ACS)

Public Use Microdata Samples (PUMS)

February 17, 2016

Marisa Hotchkiss

American Community Survey Office

Outline

- **PUMS Overview**
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

Why Use PUMS?

SEX BY SCHOOL ENROLLMENT BY EDUCATIONAL ATTAINMENT BY EMPLOYMENT STATUS FOR THE POPULATION 16 TO 19 YEARS
Universe: Population 16 to 19 years
2010-2014 American Community Survey 5-Year Estimates

- Data needed are not supported by standard tables
  - Example: narrower categories ("16 to 17 years" "18 years")
- Users want to conduct sophisticated statistical analysis
  - Example: correlation analysis (age, employment, and language)
- Can create new measures
  - Example: looking at person and household variables together (new variable: presence of working teenager in a household)

What are PUMS files?

Public **U**se **M**icrodata **S**ample

Anonymized

Downloadable

- SAS and CSV
- AFF, FTP, DataFerrett

Individual Responses

- Must be tabulated and weighted by user

Representative Sample of the Population

- 1-year (1%)
- 5-year (5%)

Summary Data vs. Microdata

What's the Difference?

	Maryland	
	Estimate	Margin of Error
Total:	5,887,776	*****
Male:	2,851,238	+/-544
Under 18 years:	687,884	+/-514
Native	657,483	+/-1,504
Foreign born:	30,401	+/-1,362
Naturalized U.S. citizen	7,636	+/-691
Not a U.S. citizen	22,765	+/-1,131
18 years and over:	2,163,354	+/-385
Native	1,786,172	+/-3,487
Foreign born:	377,182	+/-3,416
Naturalized U.S. citizen	174,573	+/-2,242
Not a U.S. citizen	202,609	+/-3,246

Aggregated tables for a geography:
"In MD, 174,573 adult males are naturalized citizens."

Individual responses:
"This 58-year-old adult male in MD is a naturalized citizen."

RT	SERIALNO	SPORDER	AGEP	CIT	SEX
P	2010000000005	1	37	1	1
P	2010000000005	2	36	1	2
P	2010000000005	3	9	1	1
P	2010000000005	4	7	1	1
P	2010000000005	5	0	1	2
P	2010000000007	1	58	4	1
P	2010000000007	2	17	1	1
P	2010000000007	3	13	1	2
P	2010000000007	4	47	4	1
P	2010000000007	5	18	1	2

Summary Data vs. Microdata

Pros and Cons

	Summary Data (ex: AFF)	Microdata (ex: PUMS)
Format	Published tables designed by Census analysts (weighted with margins of error)	Sample of household and individual responses to questionnaire (and some recoded variables)
Benefits	Easy to use Small geographies	Enables custom tables and analyses More variable detail available
Limitations	Fixed content : • categories • years • Topic combinations	More complex to use Smaller sample (larger Margins of Error) Edits to protect privacy • top coding • broader categories No geographies smaller than PUMAs

PUMS Availability

- We release 2 new PUMS files every year
  - 1-year PUMS (example: 2014 1-year PUMS)
    - October
  - 5-year PUMS (example: 2010-2014 5-year PUMS)
    - January
- Some documentation released the week prior

Multiyear (5-year) PUMS Files

- 5-year PUMS files contain the same cases as their component 1-year files

  2010 ACS 1-year PUMS
  2011 ACS 1-year PUMS
  2012 ACS 1-year PUMS = 2010-2014 ACS 5-year PUMS
  2013 ACS 1-year PUMS
  2014 ACS 1-year PUMS

- Why do we release multiyear PUMS?
  - New weights are produced using latest population estimate "vintages"
  - Coding schemes and dollar amounts are standardized to current year

Outline

- PUMS Overview
- **PUMS Geography**
- Accessing PUMS Data
- Common Questions
- Documentation and Guidance Resources

9

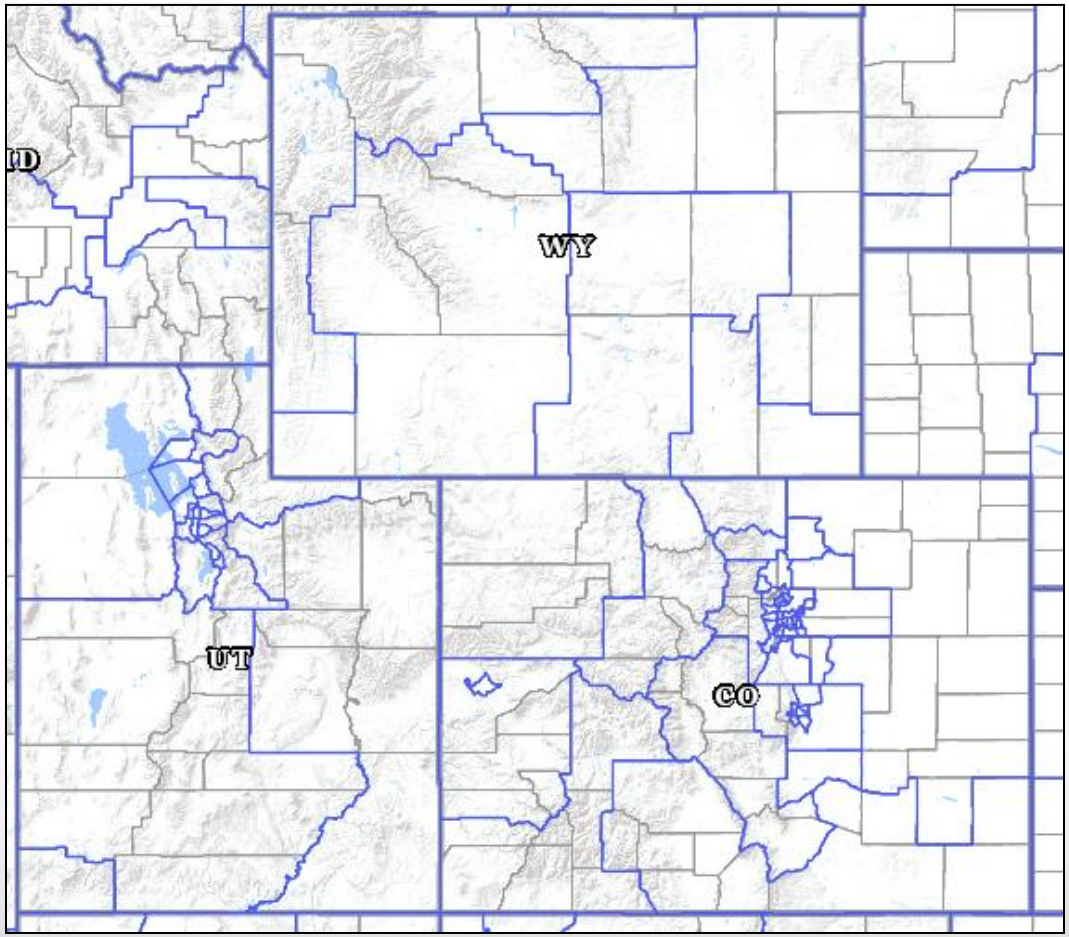
Limited Geographic Detail

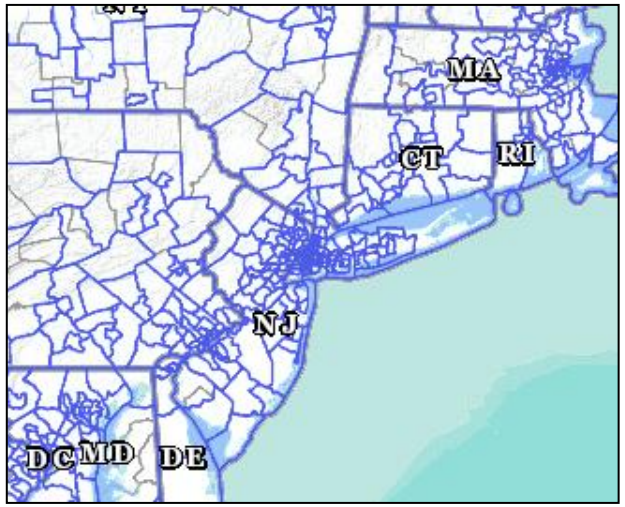
- Region, division, state, PUMA *only*
  - PUMAs can identify most cities of 100,000+ and many metropolitan areas, but not all
- PUMS is **not** designed for statistical analysis of small geographic areas

Public Use Microdata Area (PUMA)

- An area with 100,000+ population
- Identified by five-digit code (unique within each state)
- Nest within states or equivalent entities
- Geographically contiguous
- Defined after each census by states and the Census Bureau's Geography Division
  - 2010 Census PUMAs first used in the 2012 ACS
  - Multiyear files contain dual PUMA vintages

Public Use Microdata Area (PUMA) Maps





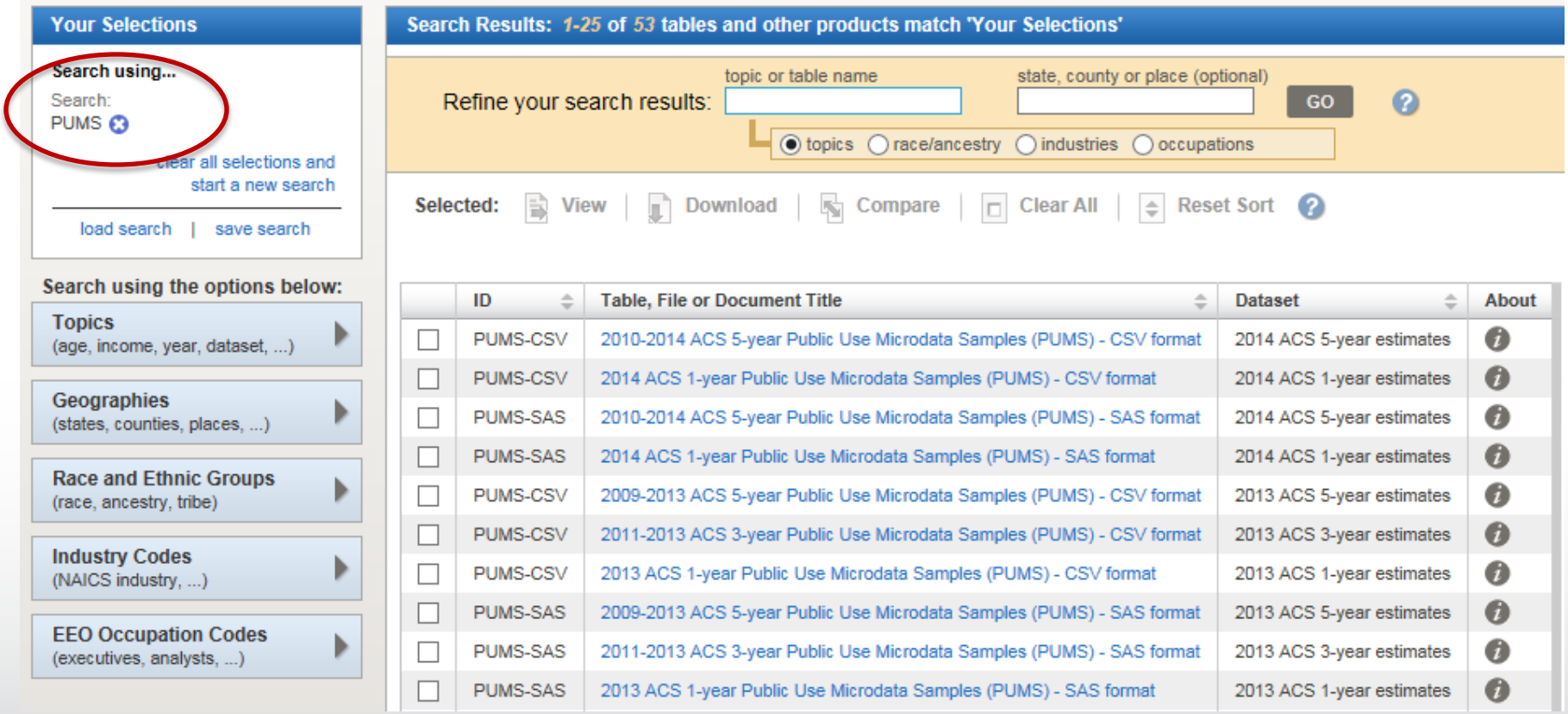
Missouri Census Data Center's MABLE, can be used to calculate the proportion of a PUMA's population in other geographies and to identify PUMA code(s) for areas of interest.
http://mcdc.missouri.edu/websas/geocorr14.html

http://www.census.gov/geo/maps-data/maps/reference.html
http://tigerweb.geo.census.gov/tigerweb/

Outline

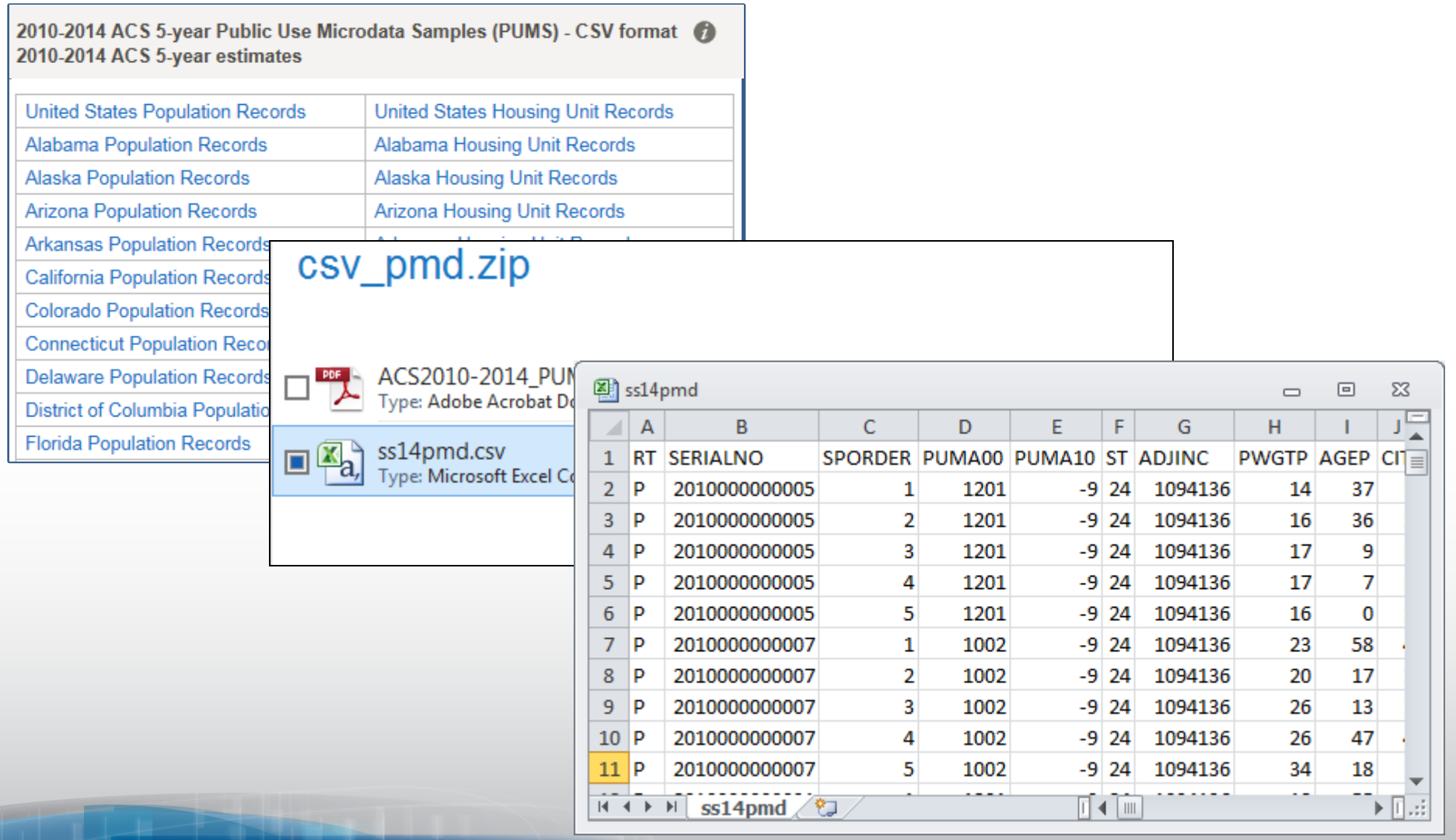
- PUMS Overview
- PUMS Geography
- **Accessing PUMS Data**
- Common User Questions
- Documentation and Guidance Resources

American FactFinder

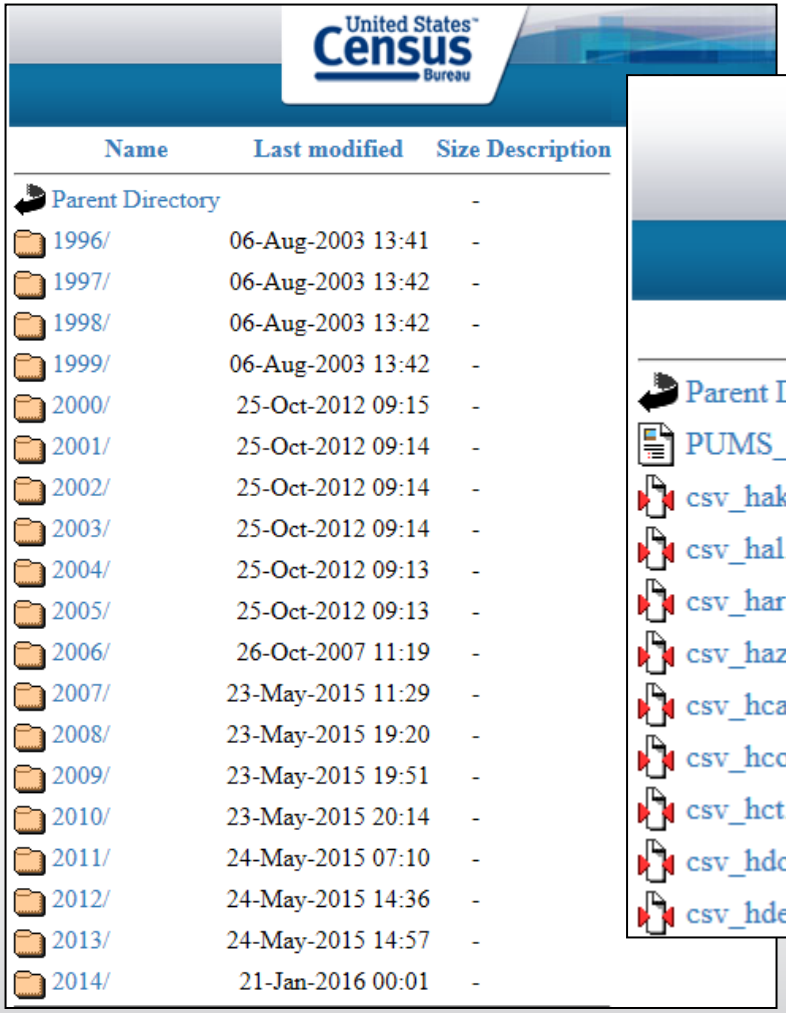


http://factfinder.census.gov

American FactFinder (cont'd)

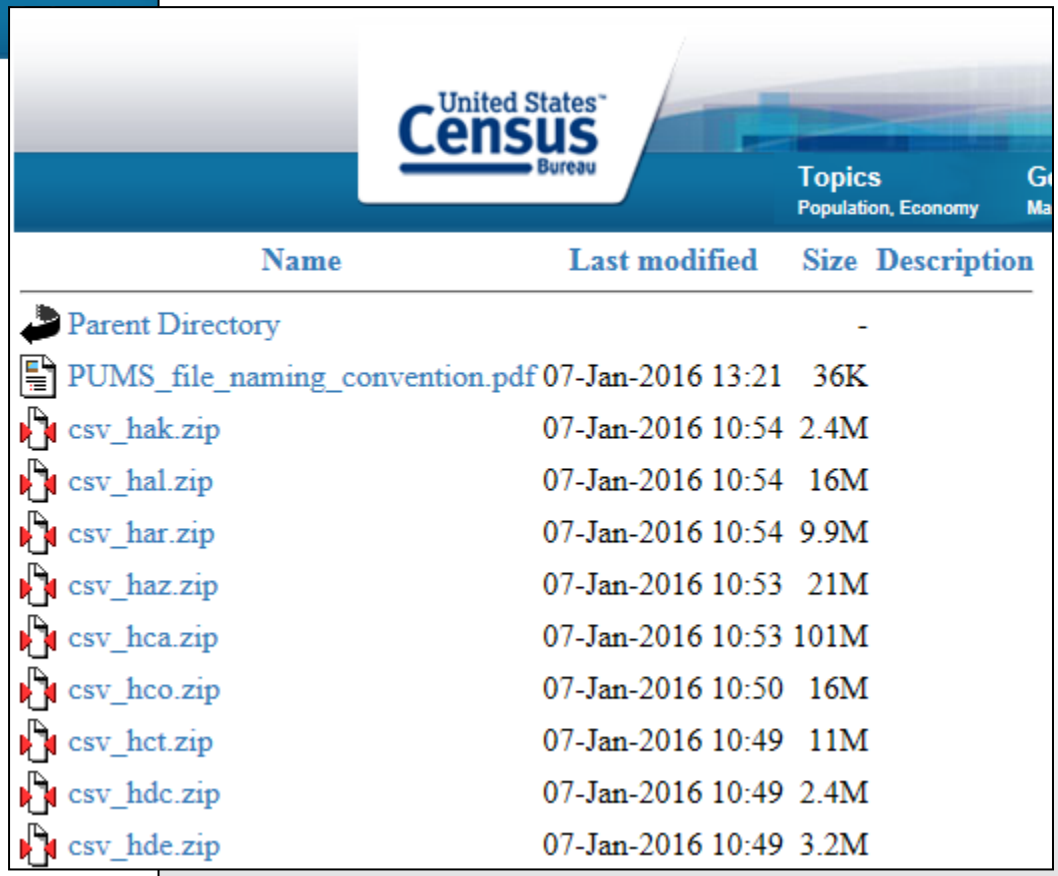


Census Bureau FTP Site



United States Census Bureau

Name	Last modified	Size	Description
Parent Directory		-	
1996/	06-Aug-2003 13:41	-	
1997/	06-Aug-2003 13:42	-	
1998/	06-Aug-2003 13:42	-	
1999/	06-Aug-2003 13:42	-	
2000/	25-Oct-2012 09:15	-	
2001/	25-Oct-2012 09:14	-	
2002/	25-Oct-2012 09:14	-	
2003/	25-Oct-2012 09:14	-	
2004/	25-Oct-2012 09:13	-	
2005/	25-Oct-2012 09:13	-	
2006/	26-Oct-2007 11:19	-	
2007/	23-May-2015 11:29	-	
2008/	23-May-2015 19:20	-	
2009/	23-May-2015 19:51	-	
2010/	23-May-2015 20:14	-	
2011/	24-May-2015 07:10	-	
2012/	24-May-2015 14:36	-	
2013/	24-May-2015 14:57	-	
2014/	21-Jan-2016 00:01	-	



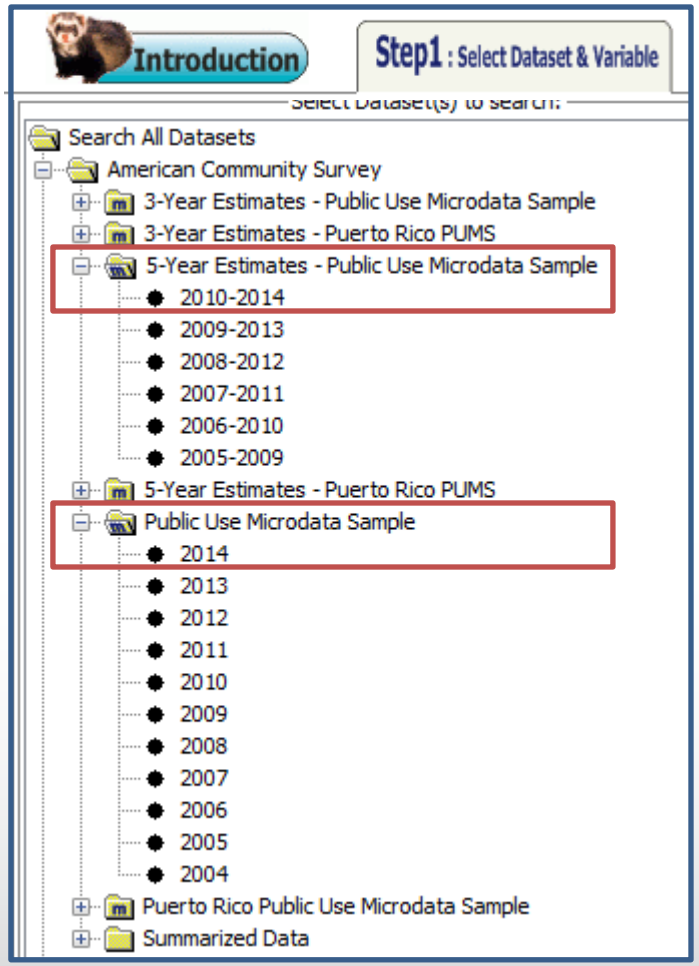
United States Census Bureau

Topics
Population, Economy

Name	Last modified	Size	Description
Parent Directory		-	
PUMS_file_naming_convention.pdf	07-Jan-2016 13:21	36K	
csv_hak.zip	07-Jan-2016 10:54	2.4M	
csv_hal.zip	07-Jan-2016 10:54	16M	
csv_har.zip	07-Jan-2016 10:54	9.9M	
csv_haz.zip	07-Jan-2016 10:53	21M	
csv_hca.zip	07-Jan-2016 10:53	101M	
csv_hco.zip	07-Jan-2016 10:50	16M	
csv_hct.zip	07-Jan-2016 10:49	11M	
csv_hdc.zip	07-Jan-2016 10:49	2.4M	
csv_hde.zip	07-Jan-2016 10:49	3.2M	

http://www2.census.gov/programs-surveys/acs/data/pums/

DataFerrett



- Menu driven system doesn't require users to have (or know how to use statistical software (i.e. SAS, STATA, SPSS, etc.)
- Can download specific variables only
- Create variable recodes

http://dataferrett.census.gov/

DataFerrett Assistance

Video Tutorials

Creating ACS Custom Tables Using DataFerrett - Part 1

May 09, 2014

Part 1 of this tutorial will show you how to make an ACS custom table using DataFerrett, the Census Bureau's free data analysis and extraction tool.

Creating ACS Custom Tables Using DataFerrett - Part 2

May 09, 2014

Part 2 of this tutorial will show you how to change the geography displayed from the default, United States, to a geography that you select.

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/dataferrett.html

What Public Use Microdata Sample (PUMS) Data Users Need to Know
(pages 12-23)

http://www.census.gov/library/publications/2009/acs/pums.html

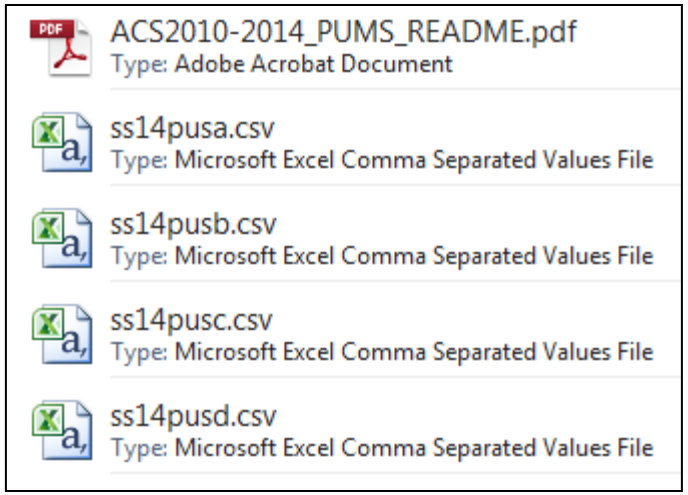
Outline

- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- **Common Questions**
- Documentation and Guidance Resources

19

How do I put PUMS files together?

- National level files must be released in pieces that users download and concatenate.
- Housing variables and person variables are separated
- Users must merge housing and person files (using SERIALNO variable as the matching variable)
- Example code is in the PUMS ReadMe and on the PUMS File Structure page




```
data population;
set psam_pusa psam_pusb;
run;

data combined; merge
population (in=pop) housing;
by serialno;
if pop;
run;
```

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/filestructure.html

Which weight should I apply?

A weight defines how many persons and housing units one PUMS sample interview represents

- WGTP: PUMS household weights
  - used to produce housing unit estimates
- PWGTP: PUMS person weights
  - used to produce population estimates
- WGTP1 – WGTP80, PWGTP1 – PWGTP80: PUMS Replicate Weights
  - used to calculate standard errors

Why don't my PUMS estimates match AFF?

Estimates generated with PUMS microdata will be slightly different from AFF estimates:

- PUMS files include only about two-thirds of the cases that were used to produce estimates on American FactFinder
- PUMS files include additional edits

When data users have doubts about their PUMS estimates, they should attempt to reproduce the PUMS Estimates for User Verification, available on the Technical Documentation page.

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

How Do I Use Dual Vintage PUMAs?

Two sets of PUMA codes on 5-year files:

- Census 2000 PUMAs
- 2010 Census PUMAs

ACS PUMS does not use overlapping geographies for any record.

Estimate of 4-yr-olds in a city (2010-2014 ACS 5-year PUMS):

4-year-olds in the 2000 PUMAs (in 2010 & 2011 records): X

4-year-olds in the 2010 PUMAs (in 2012-2014 records): Y

Add these two figures:

4-year-olds in the fuzzy PUMAs: (X+Y)

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

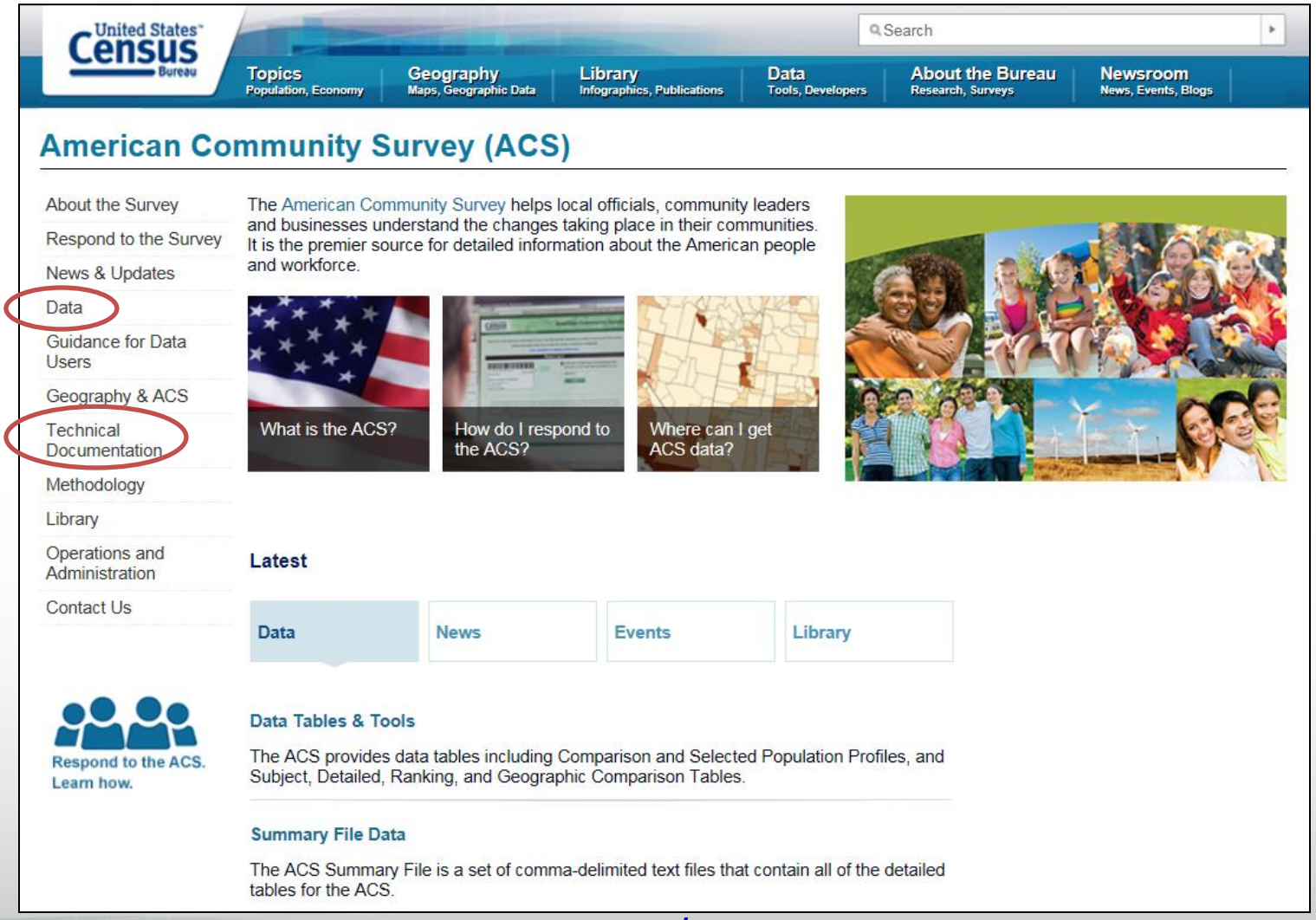
Use Caution…

- Be careful using estimates based on a small handful of cases. To obtain more cases, use multiyear files or combine geographic areas.
- Extreme values are masked to avoid disclosure. Some variables will be especially affected:
  - Dollar-amount variables (all kinds of income, mortgage, rent, utilities, property taxes, home value, property insurance costs)
  - Number of rooms and bedrooms
  - Age
  - Travel time to work, hours worked

Outline

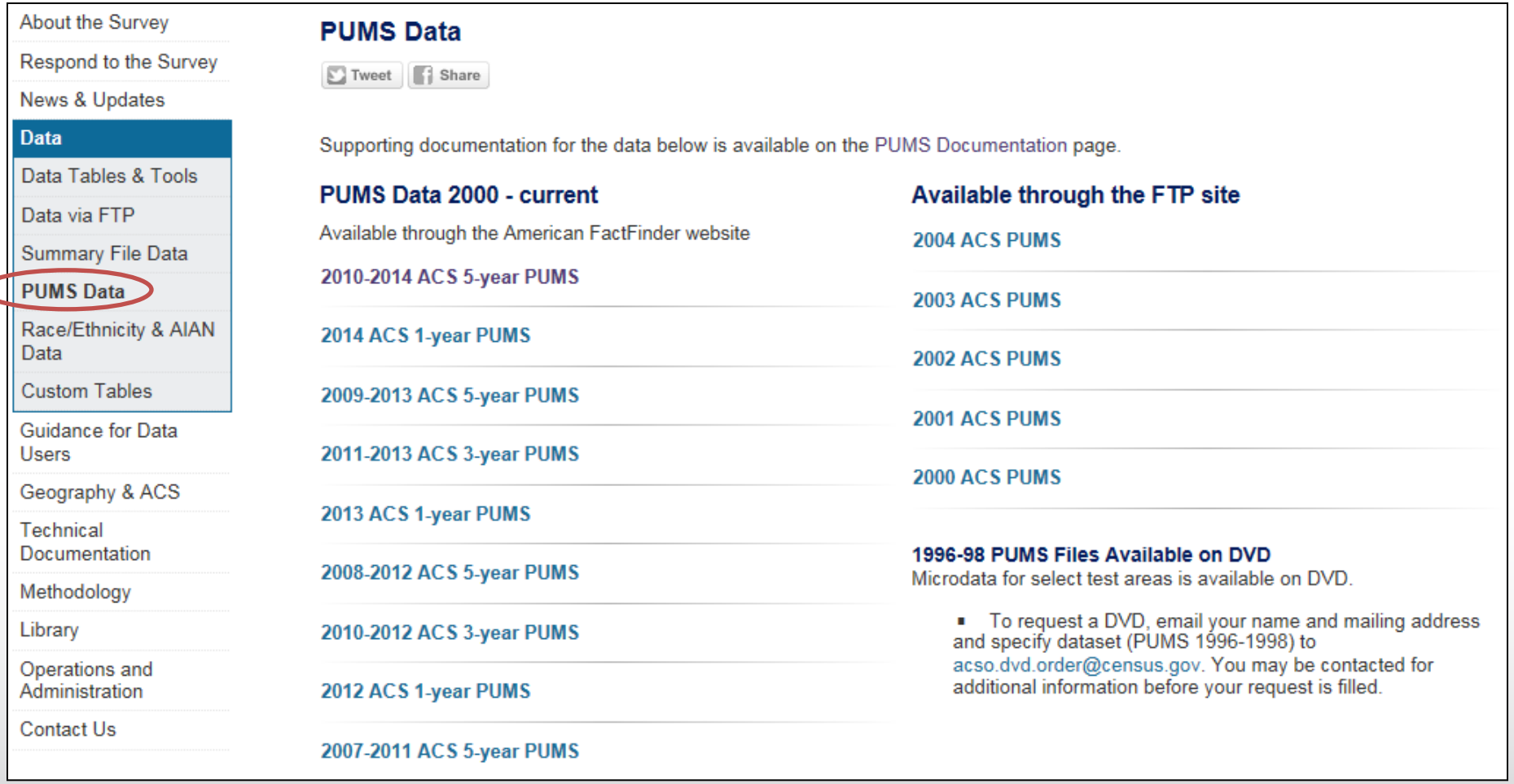
- PUMS Overview
- PUMS Geography
- Accessing PUMS Data
- Common Questions
- **Documentation and Guidance Resources**

ACS Main Page



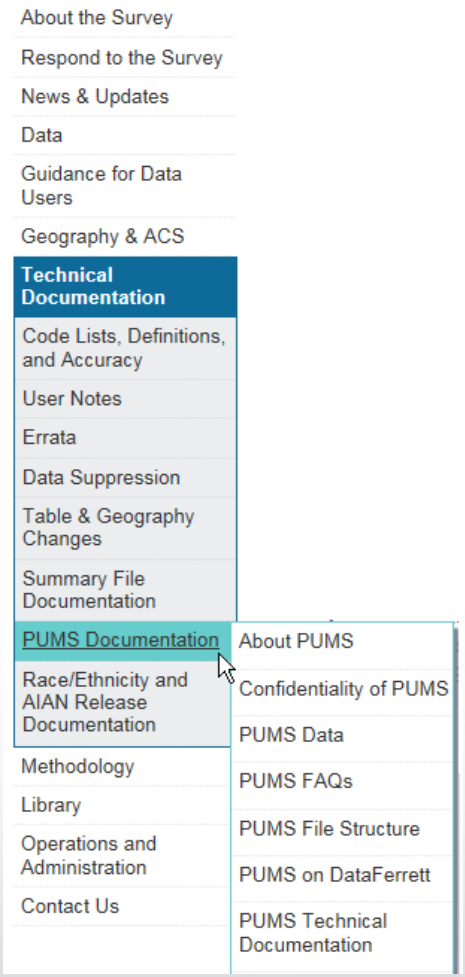
census.gov/acs

PUMS Data Page



http://www.census.gov/programs-surveys/acs/data/pums.html

PUMS Technical Documentation



- **PUMS ReadMe:** Important information about 2014 geography and variable changes as well as guidance for novice ACS PUMS files users
- **Subjects in the PUMS:** A list of topics included in each of the housing and population record files
- **PUMS Data Dictionary:** Includes variables available for each PUMS release and how each variable is coded
- **PUMS Code Lists:** Detailed codes for variables that contain a large number of coded responses, such as ancestry and occupation.
- **PUMS Top Coded and Bottom Coded Values:** List of variables with responses exceeding a state specific value that are replaced with a predetermined value
- **Accuracy of the PUMS:** A basic explanation of the sample design, estimation methodology, and accuracy of the data
- **PUMS Estimates for User Verification:** A set of Census-created PUMS estimates for data users to verify their methods

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

PUMS Data Dictionary

RT	SERIALNO	SPORDER	AGEP	CIT	SEX
P	2010000000005	1	37	1	1
P	2010000000005	2	36	1	2
P	2010000000005	3	9	1	1
P	2010000000005	4	7	1	1
P	2010000000005	5	0	1	2
P	2010000000007	1	58	4	1
P	201000				
P	201000				
P	201000				
P	201000				

```
                        PERSON RECORD—PERSON VARIABLES

AGEP            2
      Age
              00      .Under 1 year
              01..99  .1 to 99 years (Top-coded***)

CIT             1
      Citizenship status
              1  .Born in the U.S.
              2  .Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the
                 .Northern Marianas
              3  .Born abroad of American parent(s)
              4  .U.S. citizen by naturalization
              5  .Not a citizen of the U.S.
```

http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html

PUMS Resources

What Public Use Microdata Sample Data Users Need to Know

http://www.census.gov/library/publications/2009/acs/pums.html

- Overview of getting and using the data
- Outdated website instructions, but concepts are sound

ACS Public Use Microdata Samples Group on **ACS Data Users Group**

http://www.acsdatausers.org/

- Share knowledge with other PUMS users

Sign up for the ACS Public Use Microdata Sample Files alerts at https://public.govdelivery.com/accounts/USCENSUS/subscriber/new

Source Us!

U.S. Census Bureau's *[YYYY-YYYY]* American Community Survey *[1/3/5]*-year ***PUMS*** *[estimates/statistics/data release]*

PLANPHILLY a project of WHYY & newsworks.org

"As the topic of who bikes in Philly always seems to be a hot topic of conversation on social media, we reached out again for some further analysis of cycling demographics. Econsult researcher Caitlin Furio crunched the numbers for us from the **2008-2012 Public Use Microdata Sample (PUMS) data from the ACS**…"
http://www.philly.com/philly/news/Who_bikes_drives_walks_or_rides_transit_to_work_in_Philly.html

HUFFPOST BUSINESS

*"Homeownership rates are calculated from the **2013 American Community Survey (ACS) Public Use Microdata Sample (PUMS)** and are based on whether the head of household, spouse, or unmarried partner is a veteran. "*
http://www.huffingtonpost.com/jed-kolko/where-veterans-live_b_6134066.html

From *The Atlantic* CITYLAB

Household- and individual-level data are based on the. ***2012 5-Year American Community Survey (ACS) Public Use Microdata Sample (PUMS)***
http://www.citylab.com/housing/2014/08/where-private-school-enrollment-is-highest-and-lowest-across-the-us/375993/

31

Continue the Conversation #ACSdata

Sign up for and manage alerts at https://public.govdelivery.com/accounts/USCENSUS/subscriber/new

More information on the American Community Survey: http://www.census.gov/acs

(800) 923-8282 or
(301) 763-1405

facebook.com/uscensusbureau

twitter.com/uscensusbureau

youtube.com/user/uscensusbureau

instagram.com/uscensusbureau

pinterest.com/uscensusbureau

AMERICAN COMMUNITY SURVEY DATA USERS GROUP

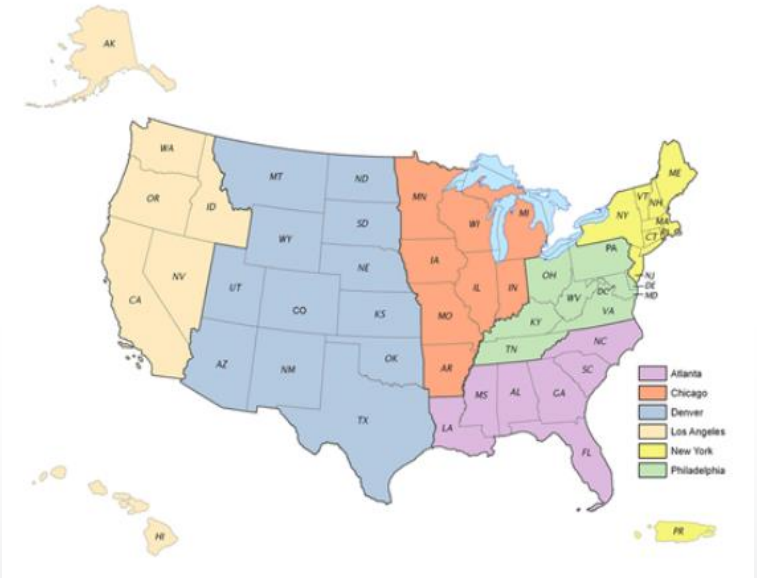
- Purpose:
  - Improve understanding of the value and utility of ACS data.
  - Promote information sharing among data users about key ACS data issues and applications
- Membership is free and open to all interested ACS data users
- Webinars and special sessions at professional meetings planned
- Users group website and online community

http://www.acsdatausers.org/

Need Local Stats?

- Assistance Near You!

  Our regional data staff can help you access local statistics from the ACS or offer training to help build your skills. Contact us at the following address: clmso.ddb.questions@census.gov



Questions?